



**University of
Zurich^{UZH}**

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2020

Pitfalls of using IQ short forms in neurodevelopmental disorders: a study in patients with congenital heart disease

Ehrler, Melanie ; Latal, Beatrice ; Polentarutti, Susanne ; von Rhein, Michael ; Held, Leonhard ; Wehrle, Flavia M

Abstract: **BACKGROUND:** Short forms of IQ (S-IQ) assessments are time efficient and highly predictive of the full IQ (F-IQ) in healthy individuals. To investigate the validity of S-IQs for patients with neurodevelopmental impairments, this study tested a well-established S-IQ version in patients with congenital heart disease (CHD). **METHODS:** The Wechsler Intelligence Scale for Children, Fourth Edition was applied in 107 children with complex CHD aged 9-11 years. F-IQ and a well-established S-IQ version were calculated for each patient. The agreement between S-IQ and F-IQ was investigated across the whole spectrum of IQ scores. Finally, we tested a method to adjust IQs to resolve potential bias and validated this method in an independent sample of 55 CHD patients. **RESULTS:** S-IQ and F-IQ correlated strongly. Nevertheless, the size of the bias correlated with the true IQ, indicating larger error at the tails of the distribution. Estimating a corrected IQ by adjusting the S-IQ with correction parameters substantially improved agreement. **CONCLUSION:** We here report that substantial bias may underestimate low IQ scores and overestimate high ones. This bias should be considered when at-risk populations are assessed with S-IQs. Importantly, the bias can be minimized by using a correction formula.

DOI: <https://doi.org/10.1038/s41390-019-0667-2>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-182249>

Journal Article

Accepted Version

Originally published at:

Ehrler, Melanie; Latal, Beatrice; Polentarutti, Susanne; von Rhein, Michael; Held, Leonhard; Wehrle, Flavia M (2020). Pitfalls of using IQ short forms in neurodevelopmental disorders: a study in patients with congenital heart disease. *Pediatric Research*, 87(5):917-923.

DOI: <https://doi.org/10.1038/s41390-019-0667-2>



Pitfalls of using IQ short forms in neurodevelopmental disorders: A study in patients with congenital heart disease

Journal:	<i>Pediatric Research</i>
Manuscript ID	PR-2019-0493.R1
Manuscript Type:	Clinical Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Ehrler, Melanie; University Children Hospital Zurich, Child Development Center; University Children Hospital Zurich, Children's Research Center Latal, Beatrice ; University Children Hospital Zurich, Child Development Center; University Children Hospital Zurich, Children's Research Center Polentarutti, Susanne; University Children Hospital Zurich, Child Development Center von Rhein, Michael; University Children Hospital Zurich, Child Development Center; University Children Hospital Zurich, Children's Research Center; Kantonsspital Winterthur, Developmental Pediatrics, Center for Social Pediatrics Held, Leonhard; University of Zurich, Department of Biostatistics Wehrle, Flavia; University Children Hospital Zurich, Child Development Center; University Children Hospital Zurich, Children's Research Center
Keywords:	Congenital Heart Disease, Neurodevelopment, Neuropsychology
Free Text Keywords:	IQ short form

SCHOLARONE™
Manuscripts

Pitfalls of using IQ short forms in neurodevelopmental disorders:
A study in patients with congenital heart disease

Melanie Ehrler^{1,2}, M.Sc., Beatrice Latal^{1,2}, MD, MPH, Susanne Polentarutti¹, M.Sc., Michael von Rhein^{1,2,3}, MD, Leonhard Held⁴, PhD, *Flavia M. Wehrle^{1,2}, PhD

Affiliations: ¹ Child Development Center, University Children’s Hospital Zurich, ZH, Switzerland; ²Children’s Research Center, University Children’s Hospital Zurich, ZH, Switzerland; ³Developmental Pediatrics, SPZ, Cantonal Hospital Winterthur, ZH, Switzerland; and ⁴Department of Biostatistics, University of Zurich, ZH, Switzerland

Address correspondence to Flavia M. Wehrle, Child Development Center, University Children’s Hospital Zurich, 75 Steinwiesstrasse, 8032 Zurich, Switzerland, [flavia.wehrle@kispi.uzh.ch].

Author Contributions: Each author has met the Pediatric research authorship requirements

Financial support: Swiss Heart Foundation, Else Kröner-Fresenius Foundation. The sponsors had no influence on study design, the collection, analysis, and interpretation of data, the writing of the report, or the decision to submit the paper for publication.

Disclosure statement: The authors have no financial relationships or conflicts of interests relevant to this article to disclose.

Category of study: Clinical research article

Author contributions:

Melanie Ehrler, M.Sc. carried out the analyses and interpretation of the data and drafted and revised the manuscript.

Prof. Dr. med. Bea Latal and Dr. med PD Michael von Rhein conceived and designed the study and critically revised the manuscript for important intellectual content.

Susanne Polentarutti, M.Sc. collected the data and critically revised the manuscript for important intellectual content.

Prof. Dr. Leonhard Held contributed substantial statistical support in order to adequately analyze the data and critically revised the manuscript for important intellectual content.

Dr. Flavia Wehrle made substantial contributions to the concept and design of the analyses and to the interpretation of the results. She further critically reviewed and revised the manuscript for important intellectual content.

All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

Content summary:

- This study investigates the agreement between IQ scores of short form and full length assessments in patients at risk for neurodevelopmental disorders
- A bias was found between full length IQs and estimated short-form IQs at the tails of the distribution
- A correction formula was provided to improve the agreement between short form and full length IQ scores
- This is of particular interest when short form IQ scores are assessed in populations with an increased proportion of individuals with low IQ scores.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Abstract

Background: Short forms of IQ (S-IQ) assessments are time-efficient and highly predictive of the full IQ (F-IQ) in healthy individuals. To investigate the validity of S-IQs for patients with neurodevelopmental impairments, this study tested a well-established S-IQ version in patients with congenital heart disease (CHD).

Methods: The Wechsler Intelligence Scale for Children, Fourth Edition was applied in 107 children with complex CHD aged 9 to 11 years. F-IQ and a well-established S-IQ version were calculated for each patient. The agreement between S-IQ and F-IQ was investigated across the whole spectrum of IQ scores. Finally, we tested a method to adjust IQs to resolve potential bias and validated this method in an independent sample of 55 CHD patients.

Results: S-IQ and F-IQ correlated strongly. Nevertheless, the size of the bias correlated with the true IQ, indicating larger error at the tails of the distribution. Estimating a corrected IQ by adjusting the S-IQ with correction parameters substantially improved agreement.

Conclusion: We here report that substantial bias may underestimate low IQ scores and overestimate high ones. This bias should be considered when at-risk populations are assessed with S-IQs. Importantly, the bias can be minimized by using a correction formula.

1. Introduction

Many neurodevelopmental disorders are characterized by deficits in general intellectual abilities, reflected in lower intelligence quotients (IQs) than those of typically developing peers. For example, the mean IQ of preterm cohorts has been reported to lie about 10 IQ points below that of typically developing children (see Allotey et al. (1) for a meta-analysis). Children with congenital heart disease (CHD) after cardiopulmonary bypass (CPB) surgery have comparable IQ deficits (2–4). Low IQ scores, particularly those below internationally recognized cut-offs defining learning and intellectual disability, are associated with a higher need for therapeutic interventions and educational action (5,6). Consequently, IQ assessments in children at risk for neurodevelopmental disorders can guide long-term care and ensure the timely initiation of support (7,8).

In research settings, IQ assessment often complements the assessment of other neurodevelopmental domains, such as executive functions (9). A common practice in such settings is to administer only a subset of the tests used for full-scale IQ (F-IQ) assessment to avoid lengthy protocols, because these may reduce performance due to fatigue or loss of motivation in patients (10,11). Regression equation models have been developed to estimate IQ from various subsets of tests (11,12). These short IQ (S-IQ) assessments have been proven to accurately predict the F-IQ in large samples of typically developing children and adolescents (10,11) and in patients with neurological disorders (13). Moreover, most children with intellectual disabilities were identified correctly when S-IQ assessments were applied (14). Importantly, the administration of S-IQ assessments has been shown to reduce testing time by more than 50% compared to the administration of the full test (15).

So far, research on the appropriateness of S-IQ assessments in neurodevelopmental disorders has been limited. Several studies have tested various S-IQ versions in patients at risk for neurodevelopmental deficits, and their findings can guide the optimal selection of subtests for

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

specific clinical populations (16–18). However, these studies have not examined whether S-IQ versions accurately measure IQ across the whole spectrum of IQ scores. In fact, there is evidence for a large bias in S-IQ scores outside the normal IQ range in healthy adults (19). This bias may be particularly pronounced in neurodevelopmental disorders due to the distribution shift towards lower IQ scores (7).

Consequently, this study aimed to comprehensively investigate the appropriateness of S-IQ assessment in a group of patients at risk for neurodevelopmental disorders, CHD patients. Specifically, we tested whether S-IQ assessments adequately estimate IQ across the whole spectrum of scores when IQ was assessed with the Wechsler Intelligence Scale for Children 4th Edition (WISC-IV). Further, we present a statistical approach to minimizing potential bias between S-IQ and F-IQ and validated this in an independent cohort of CHD patients.

2. Methods

2.1. Participants

Patients were recruited from the Research and Child Health Outcome (REACHOUT) cohort, which consists of 211 children who underwent CPB surgery at the University Children's Hospital Zurich, Switzerland between 2004 and 2009. This prospective cohort study included 5 assessments: Before surgery and at 1, 4, 6, and 10 years of age. The current study uses the data from the 10-year neurodevelopmental assessment (see Figure 1). Patients with a genetic or dysmorphic syndrome, CPB before study enrolment, or age above 6 years at first surgery were excluded from the 10-year assessment. Of 190 eligible patients, 136 agreed to participate. Twenty-eight patients did not complete all subtests of the WISC-IV required to calculate the S-IQ used in this study and were therefore excluded. In total, data of 107 patients was analyzed for the current study. To enable correction parameters to be developed, the total sample (S_1) was divided into 2 subsamples (S_{1b} , S_{1a}) by random sampling.

The reproducibility of our findings was confirmed in an independent sample of patients with CHD (S_2): We used the IQ data of 55 patients with CHD from a retrospective cohort study on CHD conducted at the University Children's Hospital Zurich. These participants underwent CPB between 1995 and 1998 and were aged 11 to 16 years at the time of the study (for further details see Schaefer et al. (4)).

The Ethical Committee of the Canton of Zurich, Switzerland approved both cohort studies. Written informed consent was obtained prior to the study participation from parents and from participants older than 11 years.

2.2. Outcome measures

IQ was assessed by well-trained professionals using the full-length German version of the WISC-IV (20). Information on the heart defect was collected from medical records. Socio-

1
2
3 economic status (SES) was estimated by rating maternal education and paternal occupation on
4
5 a 6-point scale ranging from 2 to 12 (21).
6
7 F-IQ was calculated according to the instructions provided in the manual of the WISC-IV and
8
9 adjusted for age. The S-IQ was estimated according to a regression equation method described
10
11 by Waldmann (11). We selected a subtest combination that retains the conceptual structure of
12
13 the WISC-IV to preserve construct validity. This includes one subtest from each of the 4
14
15 indices: working memory, processing speed, reasoning, and verbal comprehension. One-per-
16
17 index-subtest combinations are often used in research to estimate IQ (22–24). Of these subtest
18
19 combinations, we chose the equation with the highest reliability according to Waldmann (11)
20
21 ($R^2 = 0.911$; subtests: letter–number sequencing, symbol search, matrices, vocabulary).
22
23
24
25
26
27

28 **2.3. Statistical analyses**

29
30 Sample characteristics—sex, age, SES, and type of heart defect (cyanotic, acyanotic)—are
31
32 reported for both the full sample (S_1) and the 2 subsamples (S_{1a} , S_{1b}) and the validation sample
33
34 (S_2). Two-sided Welch’s t-test allowing for unequal variance and a two-sampled chi-squared
35
36 test were used to investigate the comparability of the 2 subsamples (S_{1a} and S_{1b}) and the full
37
38 (S_1) and validation samples (S_2).
39
40
41
42
43

44 **2.3.1. Correlation and agreement between S-IQ and F-IQ**

45
46 The correlation between the F-IQ and the S-IQ was examined using Spearman’s correlation.
47
48 We conducted 2 types of analyses to investigate the agreement between the F-IQ score and the
49
50 S-IQ. First, we examined the overall agreement: We tested for a significant mean difference
51
52 between the S-IQ and the F-IQ (i.e. the mean bias) by using a paired two-sided t-test. Further,
53
54 the proportion of participants with a bias ≤ 5 IQ-points (within 95% CI) was calculated. Second,
55
56 we analyzed whether the bias between the two IQ measurements (F-IQ, S-IQ) depends on the
57
58
59
60

IQ itself (i.e., whether the bias is equal across the whole IQ spectrum). Thus, we correlated the bias with the estimated true IQ using Spearman's correlation (25). The bias was defined as the absolute difference between the F-IQ and the S-IQ, and the true IQ was estimated from the mean of the S-IQ and F-IQ because using either value separately would cause a statistical artifact (26).

Agreement was further tested by dichotomizing the data above a clinically relevant cut-off into normal ($IQ \geq 85$) or below into abnormal IQ ($IQ < 85$). In order to test for agreement, a two-sampled chi-squared test was conducted using IQ type (F-IQ, S-IQ) as predictor and IQ score (normal, abnormal) as outcome. Further, the proportion of patients who fell into the same category for the S-IQ and the F-IQ was calculated.

2.3.2. Correction method for S-IQ.

To minimize the bias between the F-IQ and the S-IQ, we applied a statistical method to estimate a corrected IQ. Therefore, the full sample (S_1) was split into 2 subsamples by means of random sampling (S_{1b} , S_{1a}). We calculated a linear regression model for the subsample S_{1a} , using S-IQ as predictor and F-IQ as outcome ($F-IQ = \beta^0 + \beta^1 * S-IQ$). The intercept (β^0) and slope (β^1) were then used in the correction model to predict the corrected IQ of the subsample S_{1b} :

$$\text{Corrected IQ}_{1b} = \beta^0_{1a} + \beta^1_{1a} * S-IQ_{1b}$$

Two measures for quantifying the agreement with the F-IQ were calculated: the root mean square error (RMSE) and the mean absolute error (MAE). These were calculated for both the S-IQ and the corrected IQ and were then compared. A lower RMSE and MAE in the corrected IQ than in the S-IQ was interpreted as improvement of agreement.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

2.3.3. Validation of the correction method.

Finally, we validated the correction method in another independent sample of CHD patients. We estimated a corrected IQ in the validation sample (S_2) by implementing our correction parameters obtained from the subsample S_{1a} (β^0, β^1) to the S-IQ of the validation sample (S_2).

All the statistical analyses were performed using the R computing environment, version 3.5.3 (27,28). All analyses were performed with an α -level of 0.05.

3. Results

3.1. Sample characteristics

Table 1 provides sample characteristics of the full sample (S_1), the 2 subsamples (S_{1a} , S_{1b}), and the validation sample (S_2). Evidence for differences between the full and validation samples were found for age ($t(55.7) = 16.60$, $P < 0.001$, $CI-95$: from 3.03 to 3.86), and sex ($\chi^2(1) = 6.21$, $P = 0.01$). There was no difference for the type of heart defect (cyanotic and acyanotic; $\chi^2(1) = 1.64$, $P = 0.2$) and only weak evidence for a difference in SES ($t(92.9) = 1.88$, $P = 0.06$, $CI-95$: from -0.04 to 1.47). The 2 subsamples (S_{1a} , S_{1b}) did not differ in patient characteristics (see Table 1).

Table 2 provides F-IQ, S-IQ, and the mean difference between the F-IQ and the S-IQ (i.e., bias) for the full sample, the 2 subsamples and the validation sample. There was strong evidence of a difference in the mean F-IQ between the full sample S_1 and the validation sample S_2 ($t(90.6) = 3.09$, $P = 0.003$, $CI-95$: from 2.88 to 13.23). There was no evidence of a difference in mean F-IQ between the 2 subsamples (S_{1a} , S_{1b} : $t(104.0) = 0.22$, $P = 0.8$, $CI-95$: from -4.60 to 5.76).

3.2. Correlation and agreement between S-IQ and F-IQ

The analysis revealed a very strong correlation between the S-IQ and the F-IQ in the full sample (S_1 : $r = 0.95$, $P < 0.001$, Figure 2A). A two-sided paired t-test provided no evidence of a difference between S-IQ and F-IQ (S_1 : $t(106) = 1.50$, $P = 0.1$, $CI-95$: from -0.28 to 2.04). The mean difference between S-IQ and F-IQ (i.e., mean bias) in the full sample was 0.9 IQ point (Table 2). Furthermore, 81% of all participants had a bias ≤ 5 IQ points.

When considering IQ scores across the whole IQ spectrum, our findings revealed a very strong correlation between bias and the true IQ score (S_1 : $r = -0.62$, $P < 0.001$). A Bland-Altman plot

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

of difference (Figure 2B) illustrates the underestimation of S-IQ for true IQ scores below the mean and the overestimation of S-IQ for true IQ scores above the mean.

Categorizing S-IQ and F-IQ scores as normal ($\text{IQ} \geq 85$) or abnormal ($\text{IQ} < 85$) revealed no evidence for a difference in the distribution between the categories in the full sample (S_1 : $\chi^2(1) = 0.65$, $P = 0.4$). The proportion of patients of the full sample (S_1) who fell into the same category for the S-IQ and the F-IQ was 93%.

3.3. Correction method for S-IQ

The correction of the S-IQ of subsample S_{1b} with the parameters (β^0 , β^1) from the subsample S_{1a} resulted in a mean corrected IQ of 95.7 (SD = 12.2, Table 2). The correlation between the F-IQ and the corrected IQ remained strong ($r = 0.94$, $P < 0.001$, Figure 3B; for comparison: correlation between F-IQ and uncorrected S-IQ: Figure 3A). Both RMSE ($\text{RMSE}_{\text{S-IQ}} = 6.0$; $\text{RMSE}_{\text{corrected IQ}} = 4.1$) and MAE ($\text{MAE}_{\text{S-IQ}} = 4.5$; $\text{MAE}_{\text{corrected IQ}} = 3.3$) were lower in the corrected IQ than in the S-IQ. This is evidence for successful correction of the bias using the correction formula provided in this study.

To test the stability of the correction method, correction parameters (β^0 , β^1) were also extracted from subsample S_{1b} and applied to subsample S_{1a} . The results were identical (see Supplemental Figure S1 for details). Correction parameters for all one-per-index-subtest combinations with an $r > 0.9$, (11) and for further combinations consisting of 2 to 6 subtests are listed in the Supplemental Table S2a and Figure S2b.

3.3.1. Validation of the correction method

The correction method was applied in a validation sample (S_2) to further confirm its stability. Adjusting the S-IQ of the validation sample (S_2) with the correction parameters (β^0 , β^1) from the subsample (S_{1a}) resulted in a mean corrected IQ of 104.9 (SD = 15.9, Table 2). The correlation between the F-IQ and the corrected IQ remained strong ($r = 0.93$, $P < 0.001$, Supplemental Figure S3). Both RMSEs ($RMSE_{S-IQ} = 7.8$; $RMSE_{corrected\ IQ} = 4.6$) and MAE ($MAE_{S-IQ} = 6.3$; $MAE_{corrected\ IQ} = 3.4$) were lower in the corrected IQ than in the S-IQ. This is evidence for successful correction of the bias in the validation sample.

4. Discussion

In this study, we investigated whether short forms of IQ assessments (S-IQ) accurately estimate IQ scores in patients at risk for neurodevelopmental disorders. Our results indicate that while the correlation between the full-length IQ assessment (F-IQ) and S-IQ is high overall, IQ scores at the tails of the distribution are not accurately measured when using S-IQ assessments. Using our correction method, we were able to considerably reduce this bias.

In this analysis of children and adolescents with CHD, a population with frequent neurodevelopmental impairments, we calculated F-IQ scores from the full WISC-IV and S-IQ scores from a well-established 4-subtest version (11). First, we demonstrated a strong correlation between the S-IQ and the F-IQ scores. These findings are in line with previous studies on both healthy participants (10,11) and patients with neurological disorders (13,17). On examining the agreement between these two IQ measures, we observed a substantial variation in bias across the spectrum of IQ: S-IQ scores were underestimated in the lower spectrum of IQ and overestimated in the higher spectrum. This bias was replicated in various other subtest combinations (see Supplemental Table S2a and Figure S2b) and in an independent sample of CHD patients. A previous study in healthy adults that used other estimates of the Wechsler adult intelligence scale, 3rd edition, also reported larger bias at the tails of the distribution (19).

Time-efficient assessments of IQ by means of S-IQ methods are particularly valuable in research settings because IQ is often not the primary outcome of interest but serves merely as a covariate or patient characteristic to allow inferences to be drawn about potential selection bias and generalizability from the assessed cohort (29). Using short forms allows more time for the detailed assessment of developmental domains of interest, such as executive function (4,9) and memory performance (30). However, the usefulness of S-IQs has long been a matter of debate (31,32). In particular, concerns have been raised whether the construct validity

assessed in a sample with a full-length assessment holds true when only a subsample of tests is applied (11,32). For this reason, we aimed to retain the best construct validity possible by using a one-subtest-per-index short form. This approach has frequently been used in previous studies (e.g. (22–24)). Furthermore, it has been debated whether S-IQ scores are valid in samples whose IQ distributions differ from those of the general population, for example in patients with population-specific cognitive deficits. Indeed, several studies have reported considerable mean IQ differences between S-IQ and F-IQ versions in cohorts of patients with epilepsy (17), traumatic brain injury (16), neurologic disorders (13), and borderline intellectual disabilities (18). However, these studies only investigated overall mean differences between the S-IQ and the F-IQ, not the changing size and direction of bias across the whole spectrum of IQ scores. Our finding that bias is largely restricted to the tails of the distribution suggests that the significant mean differences observed in those studies likely occurred due to large biases in low IQ scores.

Importantly, the bias observed between the mean S-IQs and F-IQs of a population only provides very limited information on the agreement between the 2 measures because it does not reflect the spectrum of scores. The concept of agreement has been discussed extensively in the past. Unfortunately, there is confusion about appropriate concepts to test the agreement between 2 measurements, and this can lead to misinterpretation of results (33). Appropriate methods to investigate agreement consider the degree of bias across the whole spectrum of scores and therefore provide substantial information about the quality of a measurement (25). Therefore, this study examined bias across a whole spectrum of IQ scores by correlating the bias with the estimated true IQ, as suggested by Altman and Bland (25).

The fact that the bias occurs predominantly at the tails of the distribution may reflect a statistical phenomenon, called regression to the mean, which causes one measurement to be closer to the population mean than another measurement of the same entity. This is especially

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

apparent for values outside the normal range (34). Regression to the mean has previously been discussed as a problem affecting the interpretation of IQ scores (35–37). In particular, F-IQ scores outside the normal range tend to be closer to the mean than the corresponding S-IQ scores, since sporadic extreme subtest scores are weighted less in the F-IQ. Our findings show that it is particularly important to consider the distribution of the data when S-IQ scores of populations at risk for neurodevelopmental impairments are used in statistical analyses.

To overcome the bias between F-IQ and S-IQ measures, we tested a correction method to resolve this problem. Correction parameters were obtained from one subsample and were applied to the other subsample to estimate a corrected IQ. We observed an improved agreement between the 2 measurements: The RMSE of the S-IQ was as high as 6 before the correction, indicating low agreement (11). The RMSE of the corrected IQ was substantially reduced (RMSE = 4.1), reflecting a considerable improvement in agreement. These findings underline that correcting S-IQ scores may well improve the validity of an estimate, especially when its distribution differs from the normative sample.

The correction formula proved useful both when subdividing the full sample into 2 random subsamples and when correcting the S-IQ of an independent validation sample of CHD patients (S_2). Indeed, very similar improvements of IQ estimations were found in the validation sample that was different in regard to age, SES, and sex. The mean F-IQ of the validation sample and its distribution differed strongly from the original full sample (S_1). In fact, the mean IQ of the validation sample was above average, and its distribution was left-skewed, in all likelihood due to selection bias (4). Nevertheless, the correction formula also worked in this very different sample of patients with CHD. This implies that the correction formula we present here may also reduce bias in other samples with abnormal distributions of IQ, irrespective of the direction of the skewness. This may be of particular interest for the correction of bias in typically

developing control groups, as sampling bias is a well-known problem with high-functioning individuals (38).

If IQ estimates serve to describe the proportion of individuals with clinically relevant IQ deficits, S-IQ scores may be categorized into normal ($\text{IQ} \geq 85$) and abnormal IQ ($\text{IQ} < 85$). Our findings reveal that a high proportion of participants were categorized correctly with the S-IQ. These findings are in line with a study by Murray and colleagues (14), who used a cut-off score of 70 and successfully identified children with intellectual disabilities using an S-IQ estimation. This method is easy to apply when concerns about bias are present. However, categorizing data results in a loss of valuable information (39). Therefore, it is important to consider why the IQ has been measured and decide whether using categories only is appropriate.

Taken together, using S-IQs and the presented correction formula may be useful in research settings to provide an appropriate estimation of general intellectual abilities, especially when IQ serves as a secondary outcome. In clinical settings, the full length assessment of IQ is, however, preferable as only this provides a detailed assessment of the global IQ and the respective indices. This is required to identify deficits in specific cognitive domains and to initiate appropriate interventions.

4.1. Limitations

Our study has several limitations worth mentioning. We investigated a cohort of patients with CHD as one example of populations at risk for neurodevelopmental impairments. Therefore, our findings cannot directly be transferred to other at-risk populations. Other studies should investigate clinical populations with a higher proportion of individuals with very low IQ (e.g. severe intellectual disabilities) in order to confirm the validity of the correction formula. Also, future studies should validate our findings in populations with different profiles of cognitive impairment. Nevertheless, the cognitive profile of children born very preterm is akin to that of

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

patients with CHD (40), so we propose that the correction formula presented here could also be applied to correct the S-IQ scores of preterm-born children.

We did not have a control sample assessed with the full WISC-IV. Therefore, we could not compare the correlation and agreement of F-IQ with S-IQ between typically developing children and adolescents and those with CHD. However, a study by Spinks and colleagues (19) in healthy adults has also reported a higher proportion of large bias (> 5 IQ points) at the tails of the distribution than within the normal range. This finding indicates that the bias in the outer spectrum may also occur in cohorts of healthy individuals. The proportion of individuals with IQ scores below 85 is rather low in typically developing cohorts, though, so this may have less bearing than in at-risk populations.

This study only investigated estimates of IQ by means of equation models but did not consider other methods (e.g. Wechsler abbreviated scale for intelligence 2nd Edition (WASI-II), Wechsler 2011). This was mainly because the equation models published by Waldmann (11) are frequently used in research on German-speaking populations with various clinical characteristics (e.g. (22–24)). Further, we only investigated a variety of 4-subtest versions designed to minimize testing time while retaining the conceptual structure of the WISC-IV. Thus, we cannot draw any conclusion regarding the agreement and potential correction methods in S-IQ obtained from other estimates.

Our analyses on S-IQ were done with data of patients who conducted the full length WISC-IV assessment. Applying a short form version requires considerably less time (15), which may lead to less fatigue and better motivation (13), resulting in a better outcome. This has to be considered when short forms and the presented correction formula are used. However, the short form estimates by Waldmann (11) were also developed from data of the full length assessment. Our analyses are, therefore, in line with the original work of Waldmann.

4.2. Conclusion

S-IQs measured in a sample of children with CHD with a well-established 4-subtest version revealed a lack of agreement with F-IQs at the tails of the distribution, namely in low and high performers of the WISC-IV. This warrants particular consideration when investigating populations whose IQ scores differ from those of the general population. Bias between S-IQ and F-IQ scores can be diminished either by dichotomizing the data into scores above or below 85 or by using a correction formula presented in this study.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

5. References

1. Allotey J, Zamora J, Cheong-See F, et al. Cognitive, motor, behavioural and academic performances of children born preterm: a meta-analysis and systematic review involving 64 061 children. *BJOG An Int J Obstet Gynaecol* 2018;125:16–25.
2. Bellinger DC, Rivkin MJ, DeMaso D, et al. Adolescents with tetralogy of Fallot: neuropsychological assessment and structural brain imaging. *Cardiol Young* 2015;25:338–47.
3. Naef N, Liamlahi R, Beck I, et al. Neurodevelopmental Profiles of Children with Congenital Heart Disease at School Age. *J Pediatr* 2017;188:75–81.
4. Schaefer C, von Rhein M, Knirsch W, et al. Neurodevelopmental outcome, psychological adjustment, and quality of life in adolescents with congenital heart disease. *Dev Med Child Neurol* 2013;55:1143–9.
5. O’Brien G. Defining learning disability: what place does intelligence testing have now? *Dev Med Child Neurol* 2001;43:570–3.
6. Wieland J, Zitman FG. It is time to bring borderline intellectual functioning back into the main fold of classification systems. *BJPsych Bull* 2016;40:204–6.
7. Latal B. Neurodevelopmental Outcomes of the Child with Congenital Heart Disease. *Clin Perinatol [Internet]* 2016;43:173–85. Available from: <http://dx.doi.org/10.1016/j.clp.2015.11.012>
8. Marino BS, Lipkin PH, Newburger JW, et al. Neurodevelopmental outcomes in children with congenital heart disease: Evaluation and management. *Circulation* 2012;126:1143–72.
9. Calderon J, Bonnet D, Courtin C, Concordet S, Plumet MH, Angeard N. Executive function and theory of mind in school-aged children after neonatal corrective cardiac surgery for transposition of the great arteries. *Dev Med Child Neurol* 2010;52:1139–

- 44.
10. Crawford JR, Anderson V, Rankin PM, MacDonald J. An index-based short-form of the WISC-IV with accompanying analysis of the reliability and abnormality of differences. *Br J Clin Psychol* 2010;49:235–58.
 11. Waldmann HC. Kurzformen des HAWIK-IV: Statistische bewertung in verschiedenen anwendungsszenarien. *Diagnostica* 2008;54:202–10.
 12. Crawford JR, Allan KM, Jack AM. Short-forms of the UK WAIS—R: Regression equations and their predictive validity in a general population sample. *Br J Clin Psychol* 1992;31:191–202.
 13. van Ool JS, Hurks PPM, Snoeijen-Schouwenaars FM, et al. Accuracy of WISC-III and WAIS-IV short forms in patients with neurological disorders. *Dev Neurorehabil* 2018;21:101–7.
 14. Murray AL, McKenzie K, Murray GC. An evaluation of the performance of the WISC-IV eight-subtest short form with children who may have an intellectual disability. *J Intellect Dev Disabil* 2016;41:50–3.
 15. Ryan JJ, Glass LA, Brown CN. Administration Time Estimates for Wechsler Intelligence Scale for Children-IV Subtests, Composites, and Short Forms. *J Clin Psychol* 2007;63:309–318.
 16. Donders J, Elzinga B, Kuipers D, Helder E, Crawford JR. Development of an eight-subtest short form of the WISC-IV and evaluation of its clinical utility in children with traumatic brain injury. *Child Neuropsychol* [Internet] 2013;19:662–70. Available from: <http://dx.doi.org/10.1080/09297049.2012.723681>
 17. Hrabok M, Brooks BL, Fay-Mcclymont TB, Sherman EMS. Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV) short-form validity: A comparison study in pediatric epilepsy. *Child Neuropsychol* [Internet] 2014;20:49–59. Available from:

- <http://dx.doi.org/10.1080/09297049.2012.741225>
18. Van Duijvenbode N, Didden R, Van Den Hazel T, Engels RCME. Psychometric qualities of a tetrad WAIS-III short form for use in individuals with mild to borderline intellectual disability. *Dev Neurorehabil* 2016;19:26–30.
 19. Spinks R, Mckirgan LW, Arndt S, Caspers K, Yucuis R, Pfalzgraf CJ. IQ estimate smackdown: Comparing IQ proxy measures to the WAIS-III. *J Int Neuropsychol Soc* 2009;15:590.
 20. Peterman F, Peterman U. Hawik-IV. *Kindheit und Entwicklung* 2008;17:71–5.
 21. Largo RH, Pfister D, Molinari L, Kundu S, Lipp A DG. Significance of prenatal, perinatal and postnatal factors in the development of AGA preterm infants at five to seven years. *Dev Med Child Neurol* 1989;31:440–56.
 22. Groch S, McMakin D, Guggenbühl P, Rasch B, Huber R, Wilhelm I. Memory cueing during sleep modifies the interpretation of ambiguous scenes in adolescents and adults. *Dev Cogn Neurosci* [Internet] 2016;17:10–8. Available from: <http://dx.doi.org/10.1016/j.dcn.2015.10.006>
 23. Liechti MD, Maurizio S, Heinrich H, et al. First clinical trial of tomographic neurofeedback in attention-deficit/hyperactivity disorder: Evaluation of voluntary cortical control. *Clin Neurophysiol* [Internet] 2012;123:1989–2005. Available from: <http://dx.doi.org/10.1016/j.clinph.2012.03.016>
 24. Wehrle FM, Kaufmann L, Benz LD, et al. Very preterm adolescents show impaired performance with increasing demands in executive function tasks. *Early Hum Dev* [Internet] 2016;92:37–43. Available from: <http://dx.doi.org/10.1016/j.earlhumdev.2015.10.021>
 25. Altman D. G., Bland J. M. *Measurement in Medicine : The Analysis of Method Comparison Studies*. *Stat* 1983;32:307–17.

26. Gill S, Beevers DG, Zezulka V, Davies P. Relation between initial blood pressure and its fall with treatment. *Lancet* 1985;567–9.
27. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. 2018; Available from: <https://www.r-project.org/>
28. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. 2016;2016.
29. Steiner PM, Cook TD, Shadish WR, Clark MH. The importance of covariate selection in controlling for selection bias in observational studies. *Psychol Methods* 2010;15:250–67.
30. Cassidy AR, Newburger JW, Bellinger DC. Learning and Memory in Adolescents With Critical Biventricular Congenital Heart Disease. *J Int Neuropsychol Soc* 2017;23:627–39.
31. Donders J. Using a Short Form of the WISC-III: Sinful or Smart? *Child Neuropsychol* 2001;7:99–103.
32. Kaufman JC, Kaufman AS. Time for the changing of the guard: A farewell to short forms of intelligence tests. *J Psychoeduc Assess* 2001;19:245–67.
33. Kottner J, Streiner DL. The difference between reliability and agreement. *J Clin Epidemiol* [Internet] 2011;64:701–2. Available from: <http://dx.doi.org/10.1016/j.jclinepi.2010.12.001>
34. Bland MJ, Altman DG. Regression towards the mean. *Stat Notes* 1994;30:1499.
35. Basso MR, Bornstein RA, Roper BL, McCoy VL. Limited Accuracy of Premorbid Intelligence Estimators: A Demonstration of Regression to the Mean. *Clin Neuropsychol* 2000;14:325–40.
36. Larrabee GJ. FORUM Association Between IQ and Neuropsychological Test Performance: Commentary on Tremont, Hoffman, Scott and Adams (1998). *Clin Neuropsychol* 2000;14:139–45.

37. Veiel HOF, Koopman RF. The Bias in Regression- Based Indices of Premorbid IQ. Psychol Assess 2001;13:356–68.

38. Russell EW. Norming subjects for the Halstead Reitan battery. Arch Clin Neuropsychol 2005;20:479–84.

39. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: A bad idea. Stat Med 2006;25:127–41.

40. Easson K, Dahan-Oliel N, Rohlicek C, et al. A Comparison of Developmental Outcomes of Adolescent Neonatal Intensive Care Unit Survivors Born with a Congenital Heart Defect or Born Preterm. J Pediatr [Internet] 2018;Available from: <https://doi.org/10.1016/j.jpeds.2018.11.002>

Acknowledgement

We thank the children and parents for their participation. We thank Prof. Dr. med. Walter Knirsch for his help regarding cardiac variables.

Figure Legends

Figure 1: Flow diagram of participant selection for sample S₁. y FU = year follow up

Figure 2: A) Correlation of S-IQ and F-IQ. B) Correlation of bias and estimated true IQ, both in the full sample (S₁) (25). Dots = participants. Black line = linear regression line. Grey line = perfect agreement. Grey shadow = 95%-CI. r = Spearman’s correlation. *** = $P < 0.001$.

Figure 3: A) Correlation between F-IQ and S-IQ and B) F-IQ and *corrected* IQ in the subsample S_{1b}. Dots = participants. Black line = linear regression line. Grey line = perfect agreement. Grey shadow = 95%-CI. r = Spearman’s correlation. *** = $P < 0.001$.

Pitfalls of using IQ short forms in neurodevelopmental disorders:

A ~~validation~~ study in patients with congenital heart disease

Melanie Ehrler^{1,2}, M.Sc., Beatrice Latal^{1,2}, MD, MPH, Susanne Polentarutti¹, M.Sc., Michael von Rhein^{1,2,3}, MD, Leonhard Held⁴, PhD, *Flavia M. Wehrle^{1,2}, PhD

Affiliations: ¹ Child Development Center, University Children's Hospital Zurich, ZH, Switzerland; ²Children's Research Center, University Children's Hospital Zurich, ZH, Switzerland; ³Developmental Pediatrics, SPZ, Cantonal Hospital Winterthur, ZH, Switzerland; and ⁴Department of Biostatistics, University of Zurich, ZH, Switzerland

Address correspondence to Flavia M. Wehrle, Child Development Center, University Children's Hospital Zurich, 75 Steinwiesstrasse, 8032 Zurich, Switzerland, [flavia.wehrle@kispi.uzh.ch].

Author Contributions: Each author has met the Pediatric research authorship requirements

Financial support: Swiss Heart Foundation, Else Kröner-Fresenius Foundation. The sponsors had no influence on study design, the collection, analysis, and interpretation of data, the writing of the report, or the decision to submit the paper for publication.

Disclosure statement: The authors have no financial relationships or conflicts of interests relevant to this article to disclose.

Category of study: Clinical research article

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Author contributions:

Melanie Ehrler, M.Sc. carried out the analyses and interpretation of the data and drafted and revised the manuscript.

Prof. Dr. med. Bea Latal and Dr. med PD Michael von Rhein conceived and designed the study and critically revised the manuscript for important intellectual content.

Susanne Polentarutti, M.Sc. collected the data and critically revised the manuscript for important intellectual content.

Prof. Dr. Leonhard Held contributed substantial statistical support in order to adequately analyze the data and critically revised the manuscript for important intellectual content.

Dr. Flavia Wehrle made substantial contributions to the concept and design of the analyses and to the interpretation of the results. She further critically reviewed and revised the manuscript for important intellectual content.

All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

Content summary:

- This study investigates the agreement between IQ scores of short form and full length assessments in patients at risk for neurodevelopmental disorders
- A bias was found between full length IQs and estimated short-form IQs at the tails of the distribution
- A correction formula was provided to improve the agreement between short form and full length IQ scores
- This is of particular interest when short form IQ scores are assessed in populations with an increased proportion of individuals with low IQ scores.

Abstract

Background: Short forms of IQ (S-IQ) assessments are time-efficient and highly predictive of the full IQ (F-IQ) in healthy individuals. To investigate the validity of S-IQs for patients with neurodevelopmental impairments, this study ~~validated~~tested a well-established S-IQ version in patients with congenital heart disease (CHD).

Methods: The Wechsler Intelligence Scale for Children, Fourth Edition was applied in 107 children with complex CHD aged 9 to 11 years. F-IQ and a well-established S-IQ version were calculated for each patient. The agreement between S-IQ and F-IQ was investigated across the whole spectrum of IQ scores. Finally, we tested a method to adjust IQs to resolve potential bias and validated this method in an independent sample of 55 CHD patients.

Results: S-IQ and F-IQ correlated strongly. Nevertheless, the size of the bias correlated with the true IQ, indicating larger error at the tails of the distribution. Estimating a corrected IQ by adjusting the S-IQ with correction parameters substantially improved agreement.

Conclusion: We here report that substantial bias may underestimate low IQ scores and overestimate ~~of~~ high ones. This bias should be considered when at-risk populations are assessed with S-IQs. Importantly, the bias can be minimized by using a correction formula.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1. Introduction

Many neurodevelopmental disorders are characterized by deficits in general intellectual abilities, reflected in lower intelligence quotients (IQs) than those of typically developing peers. For example, the mean IQ of preterm cohorts has been reported to lie about 10 IQ points below that of typically developing children (see Allotey et al. (1) for a meta-analysis). Children with congenital heart disease (CHD) after cardiopulmonary bypass (CPB) surgery have comparable IQ deficits (2–4). Low IQ scores, particularly those below internationally recognized cut-offs defining learning and intellectual disability, are associated with a higher need for therapeutic interventions and educational action (5,6). Consequently, IQ assessments in children at risk for neurodevelopmental disorders can guide long-term care and ensure the timely initiation of support (7,8).

In research settings, IQ assessment often complements the assessment of other neurodevelopmental domains, such as executive functions (9). A common practice in such settings is to administer only a subset of the tests used for full-scale IQ (F-IQ) assessment to avoid lengthy protocols, because these may reduce performance due to fatigue or loss of motivation in patients (10,11). Regression equation models have been developed to estimate IQ from various subsets of tests (11,12). These short IQ (S-IQ) assessments have been proven to accurately predict the F-IQ in large samples of typically developing children and adolescents (10,11) and in patients with neurological disorders (13). Moreover, most children with intellectual disabilities were identified correctly when S-IQ assessments were applied (14). Importantly, the administration of S-IQ assessments has been shown to reduce testing time by more than 50% compared to the administration of the full test (15).

So far, research on the appropriateness of S-IQ assessments in neurodevelopmental disorders has been limited. Several studies have tested various S-IQ versions in patients at risk for neurodevelopmental deficits, and their findings can guide the optimal selection of subtests for

specific clinical populations (16–18). However, these studies have not examined whether S-IQ versions accurately measure IQ across the whole spectrum of IQ scores. In fact, there is evidence for a large bias in S-IQ scores outside the normal IQ range in healthy adults (19). This bias may be particularly pronounced in neurodevelopmental disorders due to the distribution shift towards lower IQ scores (7).

Consequently, this study aimed to comprehensively investigate the appropriateness of S-IQ assessment in a group of patients at risk for neurodevelopmental disorders, CHD patients. Specifically, we tested whether S-IQ assessments adequately estimate IQ across the whole spectrum of scores when IQ was assessed with the Wechsler Intelligence Scale for Children 4th Edition (WISC-IV). Further, we present a statistical approach to minimizing potential bias between S-IQ and F-IQ and validated this in an independent cohort of CHD patients.

2. Methods

2.1. Participants

Patients were recruited from the Research and Child Health Outcome (REACHOUT) cohort, which consists of 211 children who underwent CPB surgery at the University Children’s Hospital Zurich, Switzerland between 2004 and 2009. This prospective cohort study included 6-5 assessments: Before surgery and at 1, 4, 6, and 10 years of age. The current study uses the data from the 10-year neurodevelopmental assessment (see Figure 1). Patients with a genetic or dysmorphic syndrome, CPB before study enrolment, or age above 6 years at first surgery were excluded from the 10-year assessment. Of 190 eligible patients, 136 agreed to participate. Twenty-eight patients did not complete all subtests of the WISC-IV required to calculate the S-IQ used in this validation-study and were therefore excluded. In total, data of 107 patients was analyzed for the current validation-study. To enable correction parameters to be developed, the total sample (S_1) was divided into 2 subsamples (S_{1b} , S_{1a}) by random sampling. The reproducibility of our findings was confirmed in an independent sample of patients with CHD (S_2): We used the IQ data of 55 patients with CHD from a retrospective cohort study on CHD conducted at the University Children’s Hospital Zurich. These participants underwent CPB between 1995 and 1998 and were aged 11 to 16 years at the time of the study (for further details see Schaefer et al. (4)). The Ethical Committee of the Canton of Zurich, Switzerland approved both cohort studies. Written informed consent was obtained prior to the study participation from parents and from participants older than 11 years.

2.2. Outcome measures

IQ was assessed by well-trained professionals using the full-length German version of the WISC-IV (20). Information on the heart defect was collected from medical records. Socio-

economic status (SES) was estimated by rating maternal education and paternal occupation on a 6-point scale ranging from 2 to 12 (21).

F-IQ was calculated according to the instructions provided in the manual of the WISC-IV and adjusted for age. The S-IQ was estimated according to a regression equation method described by Waldmann (11). We selected a subtest combination that retains the conceptual structure of the WISC-IV to preserve construct validity. This includes one subtest from each of the 4 indices: working memory, processing speed, reasoning, and verbal comprehension. One-per-index-subtest combinations are often used in research to estimate IQ (22–24). Of these subtest combinations, we chose the equation with the highest reliability according to Waldmann (11) ($R^2 = 0.911$; subtests: letter–number sequencing, symbol search, matrices, vocabulary).

2.3. *Statistical analyses*

Sample characteristics—sex, age, SES, and type of heart defect (cyanotic, acyanotic)—are reported for both the full sample (S_1) and the 2 subsamples (S_{1a} , S_{1b}) and the validation sample (S_2). Two-sided Welch’s t-test allowing for unequal variance and a two-sampled chi-squared test were used to investigate the comparability of the 2 subsamples (S_{1a} and S_{1b}) and the full (S_1) and validation samples (S_2).

2.3.1. *Correlation and agreement between S-IQ and F-IQ*

The correlation between the F-IQ and the S-IQ was examined using Spearman’s correlation. We conducted 2 types of analyses to investigate the agreement between the F-IQ score and the S-IQ. First, we examined the overall agreement: We tested for a significant mean difference between the S-IQ and the F-IQ (i.e. the mean bias) by using a paired two-sided t-test. Further, the proportion of participants with a bias ≤ 5 IQ-points (within 95% CI) was calculated. Second, we analyzed whether the bias between the two IQ measurements (F-IQ, S-IQ) depends on the

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

IQ itself (i.e., whether the bias is equal across the whole IQ spectrum). Thus, we correlated the bias with the estimated true IQ using Spearman’s correlation (25). The bias was defined as the absolute difference between the F-IQ and the S-IQ, and the true IQ was estimated from the mean of the S-IQ and F-IQ because using either value separately would cause a statistical artifact (26).

Agreement was further tested by dichotomizing the data above a clinically relevant cut-off into normal ($IQ \geq 85$) or below into abnormal IQ ($IQ < 85$). In order to test for agreement, a two-sampled chi-squared test was conducted using IQ type (F-IQ, S-IQ) as predictor and IQ score (normal, abnormal) as outcome. Further, the proportion of patients who fell into the same category for the S-IQ and the F-IQ was calculated.

2.3.2. Correction method for S-IQ.

To minimize the bias between the F-IQ and the S-IQ, we applied a statistical method to estimate a corrected IQ. Therefore, the full sample (S_1) was split into 2 subsamples by means of random sampling (S_{1b} , S_{1a}). We calculated a linear regression model for the subsample S_{1a} , using S-IQ as predictor and F-IQ as outcome ($F-IQ = \beta^0 + \beta^1 * S-IQ$). The intercept (β^0) and slope (β^1) were then used in the correction model to predict the corrected IQ of the subsample S_{1b} :

Corrected IQ_{1b} = $\beta^0_{1a} + \beta^1_{1a} * S-IQ_{1b}$

Two measures for quantifying the agreement with the F-IQ were calculated: the root mean square error (RMSE) and the mean absolute error (MAE). These were calculated for both the S-IQ and the corrected IQ and were then compared. A lower RMSE and MAE in the corrected IQ than in the S-IQ was interpreted as improvement of agreement.

2.3.3. *Validation of the correction method.*

Finally, we validated the correction method in another independent sample of CHD patients.

We estimated a corrected IQ in the validation sample (S_2) by implementing our correction parameters obtained from the subsample S_{1a} (β^0 , β^1) to the S-IQ of the validation sample (S_2).

All the statistical analyses were performed using the R computing environment, version 3.5.3 (27,28). All analyses were performed with an α -level of 0.05.

3. Results

3.1. Sample characteristics

Table 1 provides sample characteristics of the full sample (S_1), the 2 subsamples (S_{1a} , S_{1b}), and the validation sample (S_2). Evidence for differences between the full and validation samples were found for age ($t(55.7) = 16.60$, $P < 0.001$, $CI-95$: from 3.03 to 3.86), and sex ($\chi^2(1) = 6.21$, $P = 0.01$). There was no difference for the type of heart defect (cyanotic and acyanotic; $\chi^2(1) = 1.64$, $P = 0.2$) and only weak evidence for a difference in SES ($t(92.9) = 1.88$, $P = 0.06$, $CI-95$: from -0.04 to 1.47). The 2 subsamples (S_{1a} , S_{1b}) did not differ in patient characteristics (see Table 1).

Table 2 provides F-IQ, S-IQ, and the mean difference between the F-IQ and the S-IQ (i.e., bias) for the full sample, the 2 subsamples and the validation sample. There was strong evidence of a difference in the mean F-IQ between the full sample S_1 and the validation sample S_2 ($t(90.6) = 3.09$, $P = 0.003$, $CI-95$: from 2.88 to 13.23). There was no evidence of a difference in mean F-IQ between the 2 subsamples (S_{1a} , S_{1b} : $t(104.0) = 0.22$, $P = 0.8$, $CI-95$: from -4.60 to 5.76).

3.2. Correlation and agreement between S-IQ and F-IQ

The analysis revealed a very strong correlation between the S-IQ and the F-IQ in the full sample (S_1 : $r = 0.95$, $P < 0.001$, Figure 2A). A two-sided paired t-test provided no evidence of a difference between S-IQ and F-IQ (S_1 : $t(106) = 1.50$, $P = 0.1$, $CI-95$: from -0.28 to 2.04). The mean difference between S-IQ and F-IQ (i.e., mean bias) in the full sample was 0.9 IQ point (Table 2). Furthermore, 81% of all participants had a bias ≤ 5 IQ points.

When considering IQ scores across the whole IQ spectrum, our findings revealed a very strong correlation between bias and the true IQ score (S_1 : $r = -0.62$, $P < 0.001$). A Bland-Altman plot of difference (Figure 2B) illustrates the underestimation of S-IQ for true IQ scores below the

mean and the overestimation of S-IQ for true IQ scores above the mean. ~~Correction parameters (for all one-per-index-subtest combinations with an $r > 0.9$, (11)) are listed in the Supplemental Table S1a and Figure S1b.~~

Categorizing S-IQ and F-IQ scores as normal ($\text{IQ} \geq 85$) or abnormal ($\text{IQ} < 85$) revealed no evidence for a difference in the distribution between the categories in the full sample (S_1 : $\chi^2(1) = 0.65$, $P = 0.4$). The proportion of patients of the full sample (S_1) who fell into the same category for the S-IQ and the F-IQ was 93%.

3.3. Correction method for S-IQ

The correction of the S-IQ of subsample S_{1b} with the parameters (β^0 , β^1) from the subsample S_{1a} resulted in a mean corrected IQ of 95.7 (SD = 12.2, Table 2). The correlation between the F-IQ and the corrected IQ remained strong ($r = 0.94$, $P < 0.001$, Figure 3B; for comparison: correlation between F-IQ and uncorrected S-IQ: Figure 3A). Both RMSE ($\text{RMSE}_{\text{S-IQ}} = 6.0$; $\text{RMSE}_{\text{corrected IQ}} = 4.1$) and MAE ($\text{MAE}_{\text{S-IQ}} = 4.5$; $\text{MAE}_{\text{corrected IQ}} = 3.3$) were lower in the corrected IQ than in the S-IQ. This is evidence for successful correction of the bias using the correction formula provided in this study.

To test the stability of the correction method, correction parameters (β^0 , β^1) were also extracted from subsample S_{1b} and applied to subsample S_{1a} . The results were identical (see Supplemental Figure S2–S1 for details). Correction parameters (for all one-per-index-subtest combinations with an $r > 0.9$, (11)) and for further combinations consisting of 2 to 6 subtests ... are listed in the Supplemental Table S2a and Figure S2b.

3.3.1. Validation of the correction method

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

The correction method was applied in a validation sample (S_2) to further confirm its stability. Adjusting the S-IQ of the validation sample (S_2) with the correction parameters (β^0, β^1) from the subsample (S_{1a}) resulted in a mean corrected IQ of 104.9 (SD = 15.9, Table 2). The correlation between the F-IQ and the corrected IQ remained strong ($r = 0.93, P < 0.001$, Supplemental Figure S3). Both RMSEs ($RMSE_{S-IQ} = 7.8$; $RMSE_{corrected\ IQ} = 4.6$) and MAE ($MAE_{S-IQ} = 6.3$; $MAE_{corrected\ IQ} = 3.4$) were lower in the corrected IQ than in the S-IQ. This is evidence for successful correction of the bias in the validation sample.

4. Discussion

In this ~~validation~~-study, we investigated whether short forms of IQ assessments (S-IQ) accurately estimate IQ scores in patients at risk for neurodevelopmental disorders. Our results indicate that while the correlation between the full-length IQ assessment (F-IQ) and S-IQ is high overall, IQ scores at the tails of the distribution are not accurately measured when using S-IQ assessments. Using our correction method, we were able to considerably reduce this bias. In this analysis of children and adolescents with CHD, a population with frequent neurodevelopmental impairments, we calculated F-IQ scores from the full WISC-IV and S-IQ scores from a well-established 4-subtest version (11). First, we demonstrated a strong correlation between the S-IQ and the F-IQ scores. These findings are in line with previous studies on both healthy participants (10,11) and patients with neurological disorders (13,17). On examining the agreement between these two IQ measures, we observed a substantial variation in bias across the spectrum of IQ: S-IQ scores were underestimated in the lower spectrum of IQ and overestimated in the higher spectrum. [This bias was replicated in various other subtest combinations \(see Supplemental Table S2a and Figure S2b\) and in an independent sample of CHD patients.](#) A previous study in healthy adults that used other estimates of the Wechsler adult intelligence scale, 3rd edition, also reported larger bias at the tails of the distribution (19).

Time-efficient assessments of IQ by means of S-IQ methods are particularly valuable in research settings because IQ is often not the primary outcome of interest but serves merely as a covariate or patient characteristic to allow inferences to be drawn about potential selection bias and generalizability from the assessed cohort (29). Using short forms allows more time for the detailed assessment of developmental domains of interest, such as executive function (4,9) and memory performance (30). However, the usefulness of S-IQs has long been a matter of debate (31,32). In particular, concerns have been raised whether the construct validity

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

assessed in a sample with a full-length assessment holds true when only a subsample of tests is applied (11,32). For this reason, we aimed to retain the best construct validity possible by using a one-subtest-per-index short form. This approach has frequently been used in previous studies (e.g. (22–24)). Furthermore, it has been debated whether S-IQ scores are valid in samples whose IQ distributions differ from those of the general population, for example in patients with population-specific cognitive deficits. Indeed, several studies have reported considerable mean IQ differences between S-IQ and F-IQ versions in cohorts of patients with epilepsy (17), traumatic brain injury (16), neurologic disorders (13), and borderline intellectual disabilities (18). However, these studies only investigated overall mean differences between the S-IQ and the F-IQ, not the changing size and direction of bias across the whole spectrum of IQ scores. Our finding that bias is largely restricted to the tails of the distribution suggests that the significant mean differences observed in those studies likely occurred due to large biases in low IQ scores.

Importantly, the bias observed between the mean S-IQs and F-IQs of a population only provides very limited information on the agreement between the 2 measures because it does not reflect the spectrum of scores. The concept of agreement has been discussed extensively in the past. Unfortunately, there is confusion about appropriate concepts to test the agreement between 2 measurements, and this can lead to misinterpretation of results (33). Appropriate methods to investigate agreement consider the degree of bias across the whole spectrum of scores and therefore provide substantial information about the quality of a measurement (25). Therefore, this ~~validation~~ study examined bias across a whole spectrum of IQ scores by correlating the bias with the estimated true IQ, as suggested by Altman and Bland (25).

The fact that the bias occurs predominantly at the tails of the distribution may reflect a statistical phenomenon, called regression to the mean, which causes one measurement to be closer to the population mean than another measurement of the same entity. This is especially

apparent for values outside the normal range (34). Regression to the mean has previously been discussed as a problem affecting the interpretation of IQ scores (35–37). In particular, F-IQ scores outside the normal range tend to be closer to the mean than the corresponding S-IQ scores, since sporadic extreme subtest scores are weighted less in the F-IQ. Our findings show that it is particularly important to consider the distribution of the data when S-IQ scores of populations at risk for neurodevelopmental impairments are used in statistical analyses.

To overcome the bias between F-IQ and S-IQ measures, we tested a correction method to resolve this problem. Correction parameters were obtained from one subsample and were applied to the other subsample to estimate a corrected IQ. We observed an improved agreement between the 2 measurements: The RMSE of the S-IQ was as high as 6 before the correction, indicating low agreement (11). The RMSE of the corrected IQ was substantially reduced (RMSE = 4.1), reflecting a considerable improvement in agreement. These findings underline that correcting S-IQ scores may well improve the validity of an estimate, especially when its distribution differs from the normative sample.

The correction formula proved useful both when subdividing the full sample into 2 random subsamples and when correcting the S-IQ of an independent validation sample of CHD patients (S_2). Indeed, very similar improvements of IQ estimations were found in the validation sample (S_2) that was different in regard to age, SES, and sex. The mean F-IQ of the validation sample and its distribution differed strongly from the original full sample (S_1). In fact, the mean IQ of the validation sample was above average, and its distribution was left-skewed, in all likelihood due to selection bias (4). Nevertheless, the correction formula also worked in this very different sample of patients with CHD. This implies that the correction formula we present here may also reduce bias in other samples with abnormal distributions of IQ, irrespective of the direction of the skewness. This may be of particular interest for the correction of bias in typically

developing control groups, as sampling bias is a well-known problem with high-functioning individuals (38).

If IQ estimates serve to describe the proportion of individuals with clinically relevant IQ deficits, S-IQ scores may be categorized into normal ($\text{IQ} \geq 85$) and abnormal IQ ($\text{IQ} < 85$). Our findings reveal that a high proportion of participants were categorized correctly with the S-IQ. These findings are in line with a study by Murray and colleagues (14), who used a cut-off score of 70 and successfully identified children with intellectual disabilities using an S-IQ estimation. This method is easy to apply when concerns about bias are present. However, categorizing data results in a loss of valuable information (39). Therefore, it is important to consider why the IQ has been measured and decide whether using categories only is appropriate.

Taken together, using S-IQs and the presented correction formula may be useful in research settings to provide an appropriate estimation of general intellectual abilities, especially when IQ serves as a secondary outcome. In clinical settings, the full length assessment of IQ is, however, preferable as only this provides a detailed assessment of the global IQ and the respective indices. This is required to identify deficits in specific cognitive domains and to initiate appropriate interventions.

4.1. Limitations

Our ~~validation~~-study has several limitations worth mentioning. We investigated a cohort of patients with CHD as one example of populations at risk for neurodevelopmental impairments. Therefore, our findings cannot directly be transferred to other at-risk populations. Other studies should investigate clinical populations with a higher proportion of individuals with very low IQ (e.g. severe intellectual disabilities) in order to confirm the validity of the correction formula. Also, future studies should validate our findings in populations with different profiles of cognitive impairment ~~may differ widely between patient populations~~ (17).

1
2
3 ~~However, Nevertheless~~, the cognitive profile of children born very preterm is akin to that of
4
5
6 patients with CHD (40), so we propose that the correction formula presented here could also
7
8 be applied to correct the S-IQ scores of preterm-born children.
9

10 We did not have a control sample assessed with the full WISC-IV. Therefore, we could not
11
12 compare the correlation and agreement of F-IQ with S-IQ~~G~~ between typically developing
13
14 children and adolescents and those with CHD. However, a study by Spinks and colleagues (19)
15
16 in healthy adults has also reported a higher proportion of large bias (> 5 IQ points) at the tails
17
18 of the distribution than within the normal range. This finding indicates that the bias in the outer
19
20 spectrum may also occur in cohorts of healthy individuals. The proportion of individuals with
21
22 IQ scores below 85 is rather low in typically developing cohorts, though, so this may have less
23
24 bearing than in at-risk populations.
25
26
27

28 This ~~validation~~ study only investigated estimates of IQ by means of equation models but did
29
30 not consider other methods (e.g. Wechsler abbreviated scale for intelligence 2nd Edition
31
32 (WASI-II), Wechsler ~~1999~~2011). This was mainly because the equation models published by
33
34 Waldmann (11) are frequently used in research on German-speaking populations with various
35
36 clinical characteristics (e.g. (22–24)). Further, we only investigated a variety of 4-subtest
37
38 versions designed to minimize testing time while retaining the conceptual structure of the
39
40 WISC-IV. Thus, we cannot draw any conclusion regarding the agreement and potential
41
42 correction methods in S-IQ obtained from other estimates.
43
44
45

46 Our analyses on S-IQ were done with data of patients who conducted the full length WISC-IV
47
48 assessment. Applying a short form version requires considerably less time (15), which may
49
50 lead to less fatigue and better motivation (13), resulting in a better outcome. This has to be
51
52 considered when short forms and the presented correction formula are used. However, the short
53
54 form estimates by Waldmann (11) were also developed from data of the full length assessment.
55
56 Our analyses are, therefore, in line with the original work of Waldmann.
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

4.2. Conclusion

S-IQs measured in a sample of children with CHD with a well-established 4-subtest version revealed a lack of agreement with F-IQs at the tails of the distribution, namely in low and high performers of the WISC-IV, ~~especially in scores outside the normal range.~~ This warrants particular consideration when investigating populations whose IQ scores differ from those of the general population. Bias between S-IQ and F-IQ scores can be diminished either by dichotomizing the data into scores above or below 85 or by using a correction formula presented in this study.

5. References

1. Allotey J, Zamora J, Cheong-See F, et al. Cognitive, motor, behavioural and academic performances of children born preterm: a meta-analysis and systematic review involving 64 061 children. *BJOG An Int J Obstet Gynaecol* 2018;125:16–25.
2. Bellinger DC, Rivkin MJ, DeMaso D, et al. Adolescents with tetralogy of Fallot: neuropsychological assessment and structural brain imaging. *Cardiol Young* 2015;25:338–47.
3. Naef N, Liamlahi R, Beck I, et al. Neurodevelopmental Profiles of Children with Congenital Heart Disease at School Age. *J Pediatr* 2017;188:75–81.
4. Schaefer C, von Rhein M, Knirsch W, et al. Neurodevelopmental outcome, psychological adjustment, and quality of life in adolescents with congenital heart disease. *Dev Med Child Neurol* 2013;55:1143–9.
5. O'Brien G. Defining learning disability: what place does intelligence testing have now? *Dev Med Child Neurol* 2001;43:570–3.
6. Wieland J, Zitman FG. It is time to bring borderline intellectual functioning back into the main fold of classification systems. *BJPsych Bull* 2016;40:204–6.
7. Latal B. Neurodevelopmental Outcomes of the Child with Congenital Heart Disease. *Clin Perinatol [Internet]* 2016;43:173–85. Available from: <http://dx.doi.org/10.1016/j.clp.2015.11.012>
8. Marino BS, Lipkin PH, Newburger JW, et al. Neurodevelopmental outcomes in children with congenital heart disease: Evaluation and management. *Circulation* 2012;126:1143–72.
9. Calderon J, Bonnet D, Courtin C, Concordet S, Plumet MH, Angeard N. Executive function and theory of mind in school-aged children after neonatal corrective cardiac surgery for transposition of the great arteries. *Dev Med Child Neurol* 2010;52:1139–

- 44.
10. Crawford JR, Anderson V, Rankin PM, MacDonald J. An index-based short-form of the WISC-IV with accompanying analysis of the reliability and abnormality of differences. *Br J Clin Psychol* 2010;49:235–58.
 11. Waldmann HC. Kurzformen des HAWIK-IV: Statistische bewertung in verschiedenen anwendungsszenarien. *Diagnostica* 2008;54:202–10.
 12. Crawford JR, Allan KM, Jack AM. Short-forms of the UK WAIS—R: Regression equations and their predictive validity in a general population sample. *Br J Clin Psychol* 1992;31:191–202.
 13. van Ool JS, Hurks PPM, Snoeijen-Schouwenaars FM, et al. Accuracy of WISC-III and WAIS-IV short forms in patients with neurological disorders. *Dev Neurorehabil* 2018;21:101–7.
 14. Murray AL, McKenzie K, Murray GC. An evaluation of the performance of the WISC-IV eight-subtest short form with children who may have an intellectual disability. *J Intellect Dev Disabil* 2016;41:50–3.
 15. Ryan JJ, Glass LA, Brown CN. Administration Time Estimates for Wechsler Intelligence Scale for Children-IV Subtests, Composites, and Short Forms. *J Clin Psychol* 2007;63:309–318.
 16. Donders J, Elzinga B, Kuipers D, Helder E, Crawford JR. Development of an eight-subtest short form of the WISC-IV and evaluation of its clinical utility in children with traumatic brain injury. *Child Neuropsychol* [Internet] 2013;19:662–70. Available from: <http://dx.doi.org/10.1080/09297049.2012.723681>
 17. Hrabok M, Brooks BL, Fay-Mcclymont TB, Sherman EMS. Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV) short-form validity: A comparison study in pediatric epilepsy. *Child Neuropsychol* [Internet] 2014;20:49–59. Available from:

- <http://dx.doi.org/10.1080/09297049.2012.741225>
18. Van Duijvenbode N, Didden R, Van Den Hazel T, Engels RCME. Psychometric qualities of a tetrad WAIS-III short form for use in individuals with mild to borderline intellectual disability. *Dev Neurorehabil* 2016;19:26–30.
 19. Spinks R, Mckirgan LW, Arndt S, Caspers K, Yucuis R, Pfalzgraf CJ. IQ estimate smackdown: Comparing IQ proxy measures to the WAIS-III. *J Int Neuropsychol Soc* 2009;15:590.
 20. Peterman F, Peterman U. Hawik-IV. *Kindheit und Entwicklung* 2008;17:71–5.
 21. Largo RH, Pfister D, Molinari L, Kundu S, Lipp A DG. Significance of prenatal, perinatal and postnatal factors in the development of AGA preterm infants at five to seven years. *Dev Med Child Neurol* 1989;31:440–56.
 22. Groch S, McMakin D, Guggenbühl P, Rasch B, Huber R, Wilhelm I. Memory cueing during sleep modifies the interpretation of ambiguous scenes in adolescents and adults. *Dev Cogn Neurosci* [Internet] 2016;17:10–8. Available from: <http://dx.doi.org/10.1016/j.dcn.2015.10.006>
 23. Liechti MD, Maurizio S, Heinrich H, et al. First clinical trial of tomographic neurofeedback in attention-deficit/hyperactivity disorder: Evaluation of voluntary cortical control. *Clin Neurophysiol* [Internet] 2012;123:1989–2005. Available from: <http://dx.doi.org/10.1016/j.clinph.2012.03.016>
 24. Wehrle FM, Kaufmann L, Benz LD, et al. Very preterm adolescents show impaired performance with increasing demands in executive function tasks. *Early Hum Dev* [Internet] 2016;92:37–43. Available from: <http://dx.doi.org/10.1016/j.earlhumdev.2015.10.021>
 25. Altman D. G., Bland J. M. *Measurement in Medicine : The Analysis of Method Comparison Studies*. *Stat* 1983;32:307–17.

26. Gill S, Beevers DG, Zezulka V, Davies P. Relation between initial blood pressure and its fall with treatment. *Lancet* 1985;567–9.

27. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. 2018;Available from: <https://www.r-project.org/>

28. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. 2016;2016.

29. Steiner PM, Cook TD, Shadish WR, Clark MH. The importance of covariate selection in controlling for selection bias in observational studies. *Psychol Methods* 2010;15:250–67.

30. Cassidy AR, Newburger JW, Bellinger DC. Learning and Memory in Adolescents With Critical Biventricular Congenital Heart Disease. *J Int Neuropsychol Soc* 2017;23:627–39.

31. Donders J. Using a Short Form of the WISC-III: Sinful or Smart? *Child Neuropsychol* 2001;7:99–103.

32. Kaufman JC, Kaufman AS. Time for the changing of the guard: A farewell to short forms of intelligence tests. *J Psychoeduc Assess* 2001;19:245–67.

33. Kottner J, Streiner DL. The difference between reliability and agreement. *J Clin Epidemiol* [Internet] 2011;64:701–2. Available from: <http://dx.doi.org/10.1016/j.jclinepi.2010.12.001>

34. Bland MJ, Altman DG. Regression towards the mean. *Stat Notes* 1994;30:1499.

35. Basso MR, Bornstein RA, Roper BL, McCoy VL. Limited Accuracy of Premorbid Intelligence Estimators: A Demonstration of Regression to the Mean. *Clin Neuropsychol* 2000;14:325–40.

36. Larrabee GJ. FORUM Association Between IQ and Neuropsychological Test Performance: Commentary on Tremont, Hoffman, Scott and Adams (1998). *Clin Neuropsychol* 2000;14:139–45.

37. Veiel HOF, Koopman RF. The Bias in Regression- Based Indices of Premorbid IQ. Psychol Assess 2001;13:356–68.
38. Russell EW. Norming subjects for the Halstead Reitan battery. Arch Clin Neuropsychol 2005;20:479–84.
39. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: A bad idea. Stat Med 2006;25:127–41.
40. Easson K, Dahan-Oliel N, Rohlicek C, et al. A Comparison of Developmental Outcomes of Adolescent Neonatal Intensive Care Unit Survivors Born with a Congenital Heart Defect or Born Preterm. J Pediatr [Internet] 2018;Available from: <https://doi.org/10.1016/j.jpeds.2018.11.002>

Acknowledgement

We thank the children and parents for their participation. [We thank Prof. Dr. med. Walter Knirsch for his help regarding cardiac variables.](#)

Figure Legends

Figure 1: Flow diagram of participant selection for sample S_1 . y FU = year follow up

Figure 2: A) Correlation of S-IQ and F-IQ. B) Correlation of bias and estimated true IQ, both in the full sample (S_1) (25). Dots = participants. Black line = linear regression line. Grey line = perfect agreement. Grey shadow = 95%-CI. r = Spearman's correlation. *** = $P < 0.001$.

Figure 3: A) Correlation between F-IQ and S-IQ and B) F-IQ and *corrected* IQ in the subsample S_{1b} . Dots = participants. Black line = linear regression line. Grey line = perfect agreement. Grey shadow = 95%-CI. r = Spearman's correlation. *** = $P < 0.001$.

Table 1: Sample characteristics of the full sample (S₁), its subsamples (S_{1a}, S_{1b}), and the validation sample (S₂)

Characteristics	Full sample (S ₁)	Subsample (S _{1a})	Subsample (S _{1b})	Validation sample (S ₂)	p-value (S ₁ , S ₂)	p-value (S _{1a} , S _{1b})
Age (Mean, range)	10.2 (9-11)	10.2 (9-11)	10.3 (9-11)	13.7(11-16)	< 0.001	0.8
Sex (% , female)	35%	32%	37%	56%	0.01	0.7
SES (Mean, range)	8.5 (4-12)	8.5 (4 – 12)	8.5 (4 – 12)	7.8(3-12)	0.06	0.9
Cyanotic CHD (%)	36%	31%	40%	47%	0.2	0.5

Table 2: Comparison of F-IQ and S-IQ for all samples and comparison of F-IQ and corrected IQ for the subsample (S_{1b}) and the validation sample (S_2).

	Mean	Standard deviation	Range	Mean Difference ^a
Full sample (S_1)				
F-IQ	95.7	13.4	53 – 120	
S-IQ	94.8	17.2	49.2 – 126.0	- 0.9
Subsample (S_{1a})				
F-IQ	95.3	14.0	53 – 120	
S-IQ	94.6	18.1	49.2 – 123.2	- 0.7
Subsample (S_{1b})				
F-IQ	95.9	12.9	61 – 118	
S-IQ	95.0	16.4	53.3 – 126.0	- 0.9
<i>Corrected IQ</i>	95.7	12.2	64.8 – 118.6	- 0.2
Validation sample (S_2)				
F-IQ	103.7	16.7	44 - 131	
S-IQ	107.5	21.5	25 - 142	-3.8
<i>Corrected IQ</i>	104.9	15.9	43.8 – 130.6	1.2

^a Mean bias between S-IQ and F-IQ.

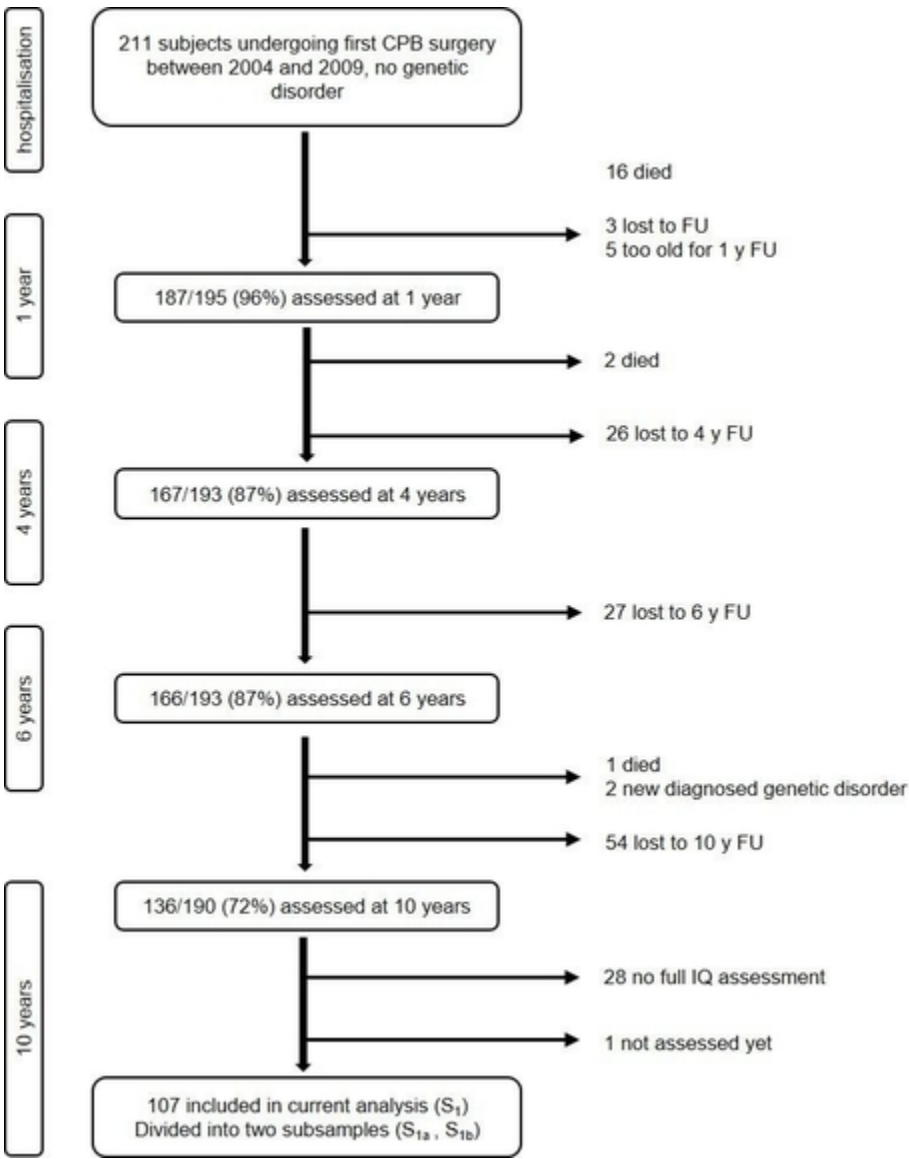


Figure 1: Flow diagram of participant selection for sample S_1 . y FU = year follow up

20x24mm (600 x 600 DPI)

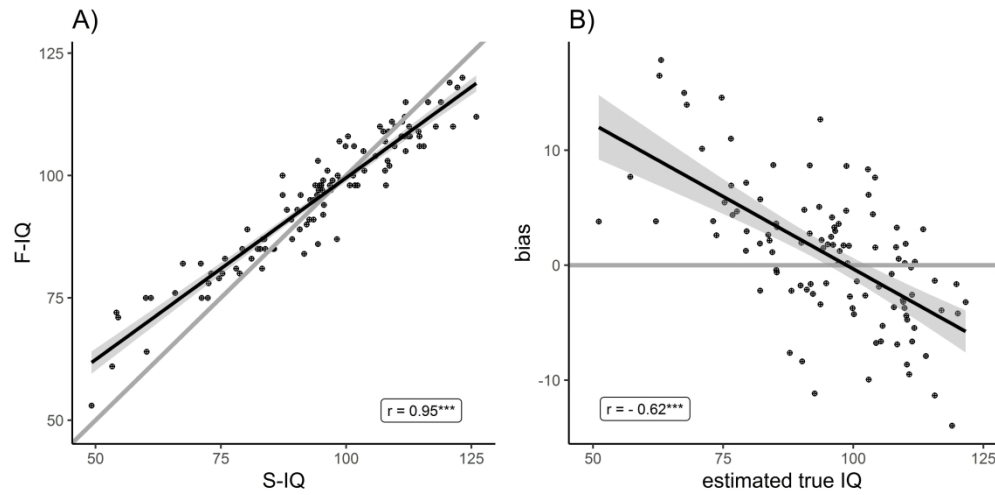


Figure 2: A) Correlation of S-IQ and F-IQ. B) Correlation of bias and estimated true IQ, both in the full sample (S_1) (25). Dots = participants. Black line = linear regression line. Grey line = perfect agreement. Grey shadow = 95%-CI. r = Spearman's correlation. *** = $P < 0.001$.

203x101mm (600 x 600 DPI)

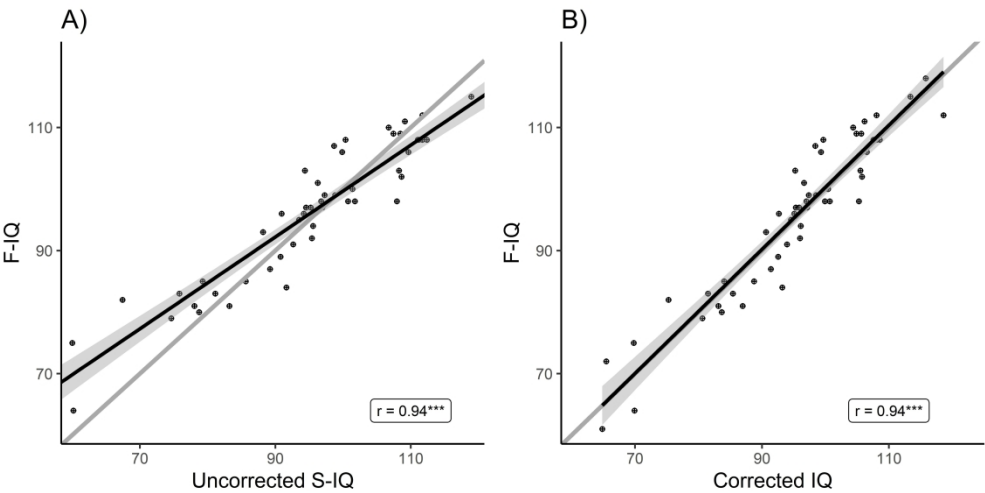
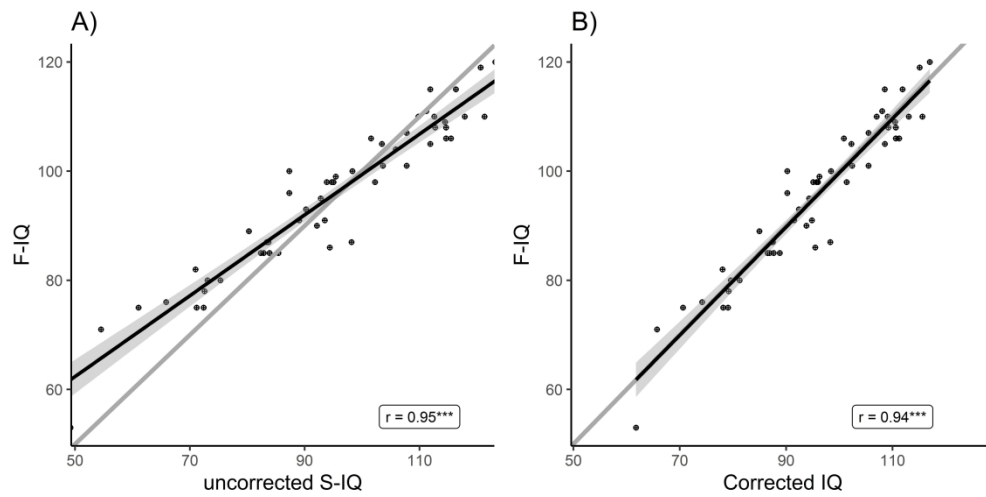


Figure 3: A) Correlation between F-IQ and S-IQ and B) F-IQ and *corrected* IQ in the subsample S_{1b} . Dots = participants. Black line = linear regression line. Grey line = perfect agreement. Grey shadow = 95%-CI. r = Spearman's correlation. *** = $P < 0.001$.

203x101mm (600 x 600 DPI)



203x101mm (600 x 600 DPI)

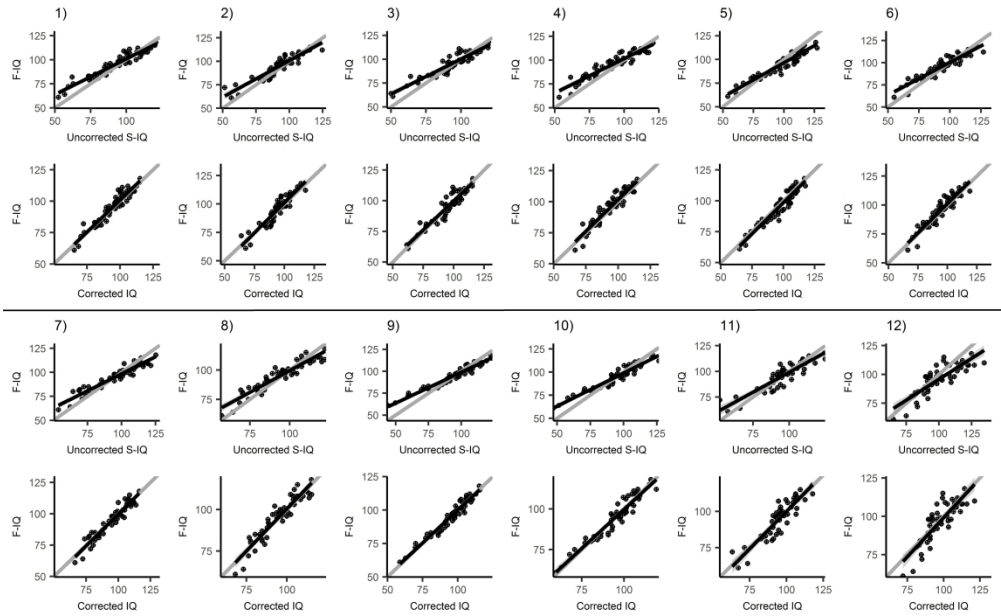
Supplemental Table S21a: Correction formula for ~~each one-per-index-subtest combination~~ with an $r > 0.9$ various subtest combinations as reported by Waldmann(11).

One-per-index- <u>S</u> _{subtest} combinations		Correction formula for corrected IQ	<u>Scenarios according to</u> <u>Waldmann (11)</u>
LNS-SS-M-V		25.366 + 0.740 * S-IQ	<u>One-per-index, Max. RSQ</u>
1	BD-V-LNS-SS	27.205 + 0.728 * S-IQ	<u>One-per-index</u>
2	DS-V-M-SS	26.162 + 0.735 * S-IQ	<u>One-per-index</u>
3	PC-V-LNS-SS	22.813 + 0.775 * S-IQ	<u>One-per-index</u>
4	BD-C-V-LNS	29.027 + 0.709 * S-IQ	<u>One-per-index</u>
5	S-LNS-M-SS	26.187 + 0.719 * S-IQ	<u>One-per-index</u>
6	C-V-LNS-M	26.258 + 0.730 * S-IQ	<u>One-per-index</u>
7	BD-S-C-LNS	28.988 + 0.697 * S-IQ	<u>One-per-index</u>
8	<u>BD-S-DS-C</u>	<u>26.445 + 0.727 * S-IQ</u>	<u>Balance</u> (M177)
9	<u>BD-PC-V-LNS-M-SS</u>	<u>26.971 + 0.720 * S-IQ</u>	<u>Max. RSQ</u> (M754)
10	<u>PC-V-LNS-M-SS</u>	<u>24.243 + 0.750 * S-IQ</u>	<u>Max. RSQ</u> (M628)
11	<u>V-M-SS</u>	<u>24.288 + 0.749 * S-IQ</u>	<u>Max. RSQ</u> (M170)
12	<u>BD-S</u>	<u>25.664 + 0.711 * S-IQ</u>	<u>Sequence</u> (M011)

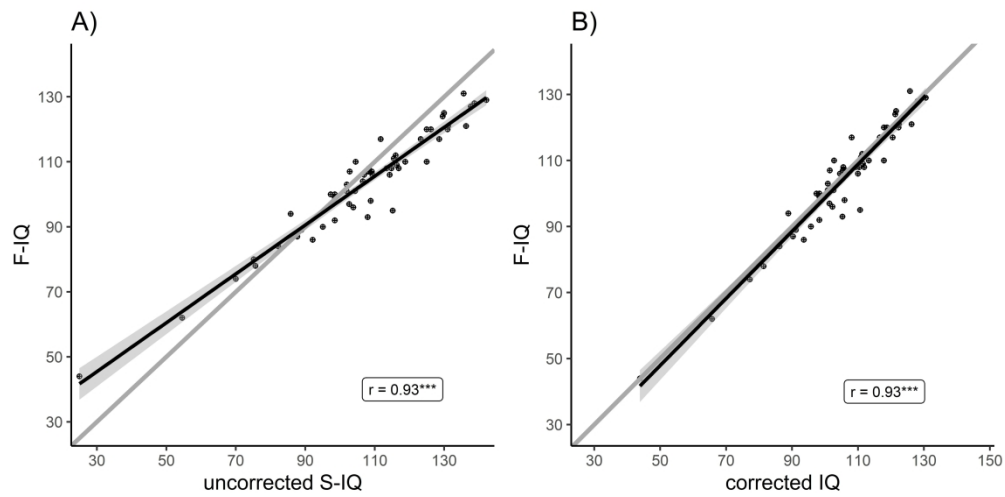
BD=Block-design, C=Coding, DS= Digit-span, LNS=letter-number-sequencing, M=Matrices,
PC=Picture-concepts, SS=symbol-search, S=Similarities, V=vocabulary

Unable to Convert Image

The dimensions of this image (in pixels) are too large to be converted. For this image to convert, the total number of pixels (height x width) must be less than 40,000,000 (40 megapixels).



203x127mm (600 x 600 DPI)



203x101mm (600 x 600 DPI)

~~Supplemental Figure S1b: Correlation between F-IQ and S-IQ, as well as F-IQ and corrected IQ in the subsample S_{1b}. Correction parameters were obtained from subsample S_{1a}. Numbers correspond to Supplemental Table S_{1a}. Dots = participants. Black line = linear regression line. Grey line = perfect agreement. Grey shadow = 95%-CI.~~

Supplemental Figure S2S1: A) Correlation between F-IQ and S-IQ and B) F-IQ and *corrected* IQ in the subsample n_{1a}. Correction parameters were obtained from subsample n_{1b}S_{1b}. Dots = participants. Black line = linear regression line. Grey line = perfect agreement. Grey shadow = 95%-CI. *r* = Spearman’s correlation. *** = $P < 0.001$.

Supplemental Figure S12b: Correlation between F-IQ and S-IQ, as well as F-IQ and *corrected* IQ in the subsample S_{1b}. Correction parameters were obtained from subsample S_{1a}. Numbers correspond to Supplemental Table S_{12a}. Dots = participants. Black line = linear regression line. Grey line = perfect agreement. Grey shadow = 95%-CI.

Supplemental Figure S3: A) Correlation between F-IQ and S-IQ and B) F-IQ and *corrected* IQ in the replication sample n₂. Correction parameters were obtained from subsample n_{1b}. Dots = participants. Black line = linear regression line. Grey line = perfect agreement. Grey shadow = 95%-CI. *r* = Spearman’s correlation. *** = $P < 0.001$.